

FILE COPY

DR InterPro: IPR000152; Asx\_hydroxyl.  
 DR InterPro: IPR000561; EGF-like.  
 DR InterPro: IPR001881; EGF\_Ca.  
 DR InterPro: IPR002174; Furin-like.  
 DR InterPro: IPR002049; Laminin\_EGF.  
 DR SMART: SM00181; EGF; 3.  
 DR SMART: SM00179; EGF\_Ca; 2.  
 DR SMART: SM00001; EGF\_Like; 1.  
 DR SMART: SM00261; FU; 2.  
 DR PROSITE: PS00010; ASX\_HYDROXYL; 1.  
 DR PROSITE: PS00022; EGF\_1; UNKNOWN\_1.  
 DR PROSITE: PS01187; EGF\_Ca; 2.  
 DR PROSITE: PS01186; EGF\_2; 2.  
 DR PROSITE: PS01248; LAMININ\_TYPE\_EGF; UNKNOWN\_1.  
 DR Calcium-binding; EGF-like domain; Glycoprotein; Hydroxylation;  
 KW Hypothetical protein; Repeat.  
 SQ SEQUENCE 321 AA; 34958 MW; A74360A1D817F23D CRC64;

Query Match 89.1%; Score 1787; DB 4; Length 321;  
 Best Local Similarity 90.9%; Pred. No. 9.1e-164;  
 Matches 321; Conservative 0; Mismatches 0; Indels 32; Gaps 1;

QY 1 MRLPRRAALGLLPLLLPAPAAKPTCHRCRGLVDKFNQGMVDTAKNFGGGNTAW 60  
 DB 1 MRLPRRAALGLLPLLLPAPAAKPTCHRCRGLVDKFNQGMVDTAKNFGGGNTAW 60  
 QY 61 EETLSKYESSEIRLLEILGLCESSDFECNOMLEAEHLEAWLQKSEYDPLFEWFC 120  
 DB 61 EETLSKYESSEIRLLEILGLCESSDFECNOMLEAEHLEAWLQKSEYDPLFEWFC 120  
 QY 121 VKTLKVCSPGTGPDCLACQGGSGRQPCSGNGHCSGDSGRQSGRCHMGYQGPCLTDC 180  
 DB 121 VKTLKVCSPGTGPDCLACQGGSGRQPCSGNGHCSGDSGRQSGRCHMGYQGPCLTDC 180  
 QY 181 MDGYFSSLRNETHSICITACDESKTCSGLNTRDGCCEVGVWLDGACVVDDECAAEPPP 240  
 DB 181 MDGYFSSLRNETHSICITACDESKTCSGLNTRDGCCEVGVWLDGACVVDDECAAEPPP 240  
 QY 241 CSAAQFCNANGSYTCRECDSSCGVCTGEGPNCCEICISGAREHGQADVDDECSLAET 300  
 DB 241 CSAAQFCNANGSYTCRECDSSCGVCTGEGPNCCEICISGAREHGQADVDDECSLAET 300  
 QY 301 CVRKNENYNTPGSYVCPDGFEEEDACVPPAEAEATESPTQLPSREDL 353  
 DB 269 CVRKNENYNTPGSYVCPDGFEEEDACVPPAEAEATESPTQLPSREDL 321

RESULT 2  
 ID Q9CYAO PRELIMINARY; PRT; 350 AA.  
 AC Q9CYAO;  
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
 DE 5730592L21RIK  
 GN 5730592L21RIK.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=EMBRYO;  
 RX MEDLINE=21085660; Pubmed=11217851;  
 RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
 RA Alizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,  
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
 RA Fleischnann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
 RA Ruelh P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,  
 RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,

DR  
 DR  
 DR

11

RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 RA Suzuki H., Sato K., Schoenbach C., Seiya T., Shibata Y., Storch K.-F.,  
 RA Suzuki H., Toyo-Oka K., Wang K.H., Weitz C., Whittaker C., Wilmig L.,  
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,  
 RA Hayashizaki Y.,  
 RT "Functional annotation of a full-length mouse cDNA collection."  
 RL Nature 409:685-690 (2001).  
 DR EMBL; AK017880; BAB30986.1;  
 DR HSSP; P35555; 1EMN.  
 DR MGD; MGI:1923987; 5730592L21RIK.  
 DR InterPro: IPR000152; Asx\_hydroxyl.  
 DR InterPro: IPR000561; EGF-Like.  
 DR InterPro: IPR001881; EGF\_Ca.  
 DR InterPro: IPR002174; Furin-like.  
 DR SMART: SM00181; EGF; 4.  
 DR SMART: SM00179; EGF\_Ca; 3.  
 DR SMART: SM00001; EGF\_Like; 1.  
 DR SMART: SM00261; FU; 3.  
 DR PROSITE: PS00010; ASX\_HYDROXYL; 1.  
 KW EGF-like domain; Glycoprotein; Hydroxylation.  
 SQ SEQUENCE 350 AA; 38219 MW; 781D7389B1944231 CRC64;

Query Match 76.5%; Score 1533.5; DB 11; Length 350;  
 Best Local Similarity 75.6%; Pred. No. 2.4e-139;  
 Matches 267; Conservative 30; Mismatches 53; Indels 3; Gaps 2;

QY 1 MRLPRRAALGLLPLLLPAPAAKPTCHRCRGLVDKFNQGMVDTAKNFGGGNTAW 60  
 DB 1 MLLLLAAEGL--LLLLPPPAGVASKRPTMCRCLTVDFKNQGMANTAKNFGGGNTAW 58  
 QY 61 EETLSKYESSEIRLLEILGLCESSDFECNOMLEAEHLEAWLQKSEYDPLFEWFC 120  
 DB 59 EETLSKYESSEIRLLEILGLCESSDFECNOMLEAEHLEAWLQKSEYDPLFEWFC 118  
 QY 121 VKTLKVCSPGTGPDCLACQGGSGRQPCSGNGHCSGDSGRQSGRCHMGYQGPCLTDC 180  
 DB 119 VHTLKACCLPGTYGPDQCQGGSERPCSGNGYCSGDSGRQSGRCHMGYQGPCLTDC 178  
 QY 181 MDGYFSSLRNETHSICITACDESKTCSGLNTRDGCCEVGVWLDGACVVDDECAAEPPP 240  
 DB 179 TDGFFSLRNETHSICITACDESKTCSGLNTRDGCCEVGVWLDGACVVDDECAAEPPP 238  
 QY 241 CSAAQFCNANGSYTCRECDSSCGVCTGEGPNCCEICISGAREHGQADVDDECSLAET 300  
 DB 239 CSDGYCENVNGSYTCEDCDSTCVCTGKGPANCKEAGYTKESGQCTDIDCSLEKA 298  
 QY 301 CVRKNENYNTPGSYVCPDGFEEEDACVPPAEAEATESPTQLPSREDL 353  
 DB 299 CVRKNENYNTPGSYVCPDGFEEEDACVPPAEAEATESPTQLPSREDL 350

RESULT 3  
 ID Q60438 PRELIMINARY; PRT; 348 AA.  
 AC Q60438;  
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
 DE HT PROTEIN.  
 OS Cricetus griseus (Chinese hamster).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;  
 OC Cricetulus.  
 OX NCBI\_TaxID=10029;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Chen H., Okubo, T., Ling V., Zhang W.;

RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.

DR EMBL; U48852; AA91469.1; -  
 DR HSP; P35555; IMN.  
 DR InterPro; IPR000152; Asx\_hydroxyl.  
 DR InterPro; IPR000561; EGF-like.  
 DR InterPro; IPR001881; EGF\_Ca.  
 DR InterPro; IPR002174; Furin-like.  
 DR InterPro; IPR002049; Laminin\_EGF.  
 DR InterPro; IPR001005; Myb\_DNA\_bind.  
 DR SMART; SM00179; EGF\_CA; 1.  
 DR SMART; SM00001; EGF\_like; 2.  
 DR SMART; SM00261; FU; 2.  
 DR PROSITE; PS00010; ASX\_HYDROXYL; 1.  
 DR PROSITE; PS00022; EGF\_1; UNKNOWN\_1.  
 DR PROSITE; PS01186; EGF\_2; 2.  
 DR PROSITE; PS01187; EGF\_CA; 2.  
 DR PROSITE; PS01248; LAMININ\_TYPE\_EGF; UNKNOWN\_1.  
 DR PROSITE; PS00037; MYB\_1; UNKNOWN\_1.  
 KW Calcium-binding; EGF-like domain; Glycoprotein; Hydroxylation; Repeat.  
 SQ SEQUENCE 348 AA; 38199 MW; BD61F6C89971BB6D CRC64;

Query Match 73.98; Score 1481.5; DB 11; Length 348;

Best Local Similarity 73.98; Pred. No. 2.4e-134;  
 Matches 255; Conservative 31; Mismatches 58; Indels 1; Gaps 1;

Qy 1 MRLPRAALGLLLPAPPAEAKPTCHRCRGLVDKFNQGVDTAKNFGGNTAW 60  
 Db 1 MHLPPAAVGLL-LLLLPPPARVASRFTWCORCALVDKFNQGMANTARNFGGNTAW 59  
 Qy 61 EEKYLKYSSEIRLEILEGLGESSDFECNMLEAEHEHLEAWLQKSEYDLPFWFC 120  
 Db 60 EEKSLKYSFEIRLEILEGLGESSDFECNMLEAEHEHLEAWLQKSEYDLPFWFC 119  
 Qy 121 VKTLKVCSPGTGPDCLACOGSGRSCGNGHSCGSDGSRGDSCHRMGYQGPLTDC 180  
 Db 120 VHTLAKCLPTGYDCCOGSGRSCGNGHSCGSDGSRGDSCHRMGYQGPLTDC 179  
 Qy 181 MDGYFSLRNETHSITCTACDESKTCSGLTNRDCGCEVGVWLDGACVDVDECAEPPP 240  
 Db 180 MDGYFSLRNETHSITCTACDESKTCSGLTNRDCGCEVGVWLDGACVDVDECAEPPP 239  
 Qy 241 CSAQAFKNGSYTCEDSCSCVGTGEGPGNCKECISGYAREHGQADVDGSLAETP 300  
 Db 240 CSNVIYCNVNGSYTCEDSCSCVGTGEGPGNCKECISGYAREHGQADVDGSLAETP 299  
 Qy 301 CVRNENYNTPGSVVCVCPDGFETEDACVPPAEATEGESPT 345  
 Db 300 CKENENYNTPGSVVCVCPDGFETEDACVPPAEATEGESPT 344

RESULT 4

Qy 096HD1 PRELIMINARY; PRT; 422 AA.  
 AC Q96HD1 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DE UNKNOWN (PROTEIN FOR MGC:8447).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=LUNG CARCINOMA;  
 RA Strausberg R.;  
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC008720; AA080720.1; -  
 SQ SEQUENCE 422 AA; 45909 MW; C26CB6EF7052E1B5 CRC64;

Query Match 48.3%; Score 969; DB 4; Length 422;

Best Local Similarity 47.1%; Pred. No. 5.8e-85;  
 Matches 168; Conservative 50; Mismatches 119; Indels 20; Gaps 4;  
 Qy 7 AALGLLLPLLLLP-----PAPEAA--KKPTCHRCRGLVDKFNQGVDTAKNFGGG 56  
 Db 12 AVLWGLSLFLNLPQPIWLQSPPPQPPQPHCPHCRGLVDKFNQGLERTIRDNFGGG 71  
 Qy 57 NTAAWEKTLKYSSEIRLEILEGLGESSDFECNMLEAEHEHLEAWLQKSEYDLPF 116  
 Db 72 NTAAWEENLSKYNDSETRLVEVLGVCSKSDFECHRLLELSEELVESWFWFKQAEPLF 131  
 Qy 117 EWFVCVTKLVCCSPGTGPDCLACOGSGRSCGNGHSCGSDGSRGDSCHRMGYQGPL 176  
 Db 132 QWLCSDSLKCCPAGTGPSCPCPGGTERPCGGYGOCEGEGTRGGSGHCDCAQYGEA 191  
 Qy 177 CTCMDCYFSSLRNETHSITCTACDESKTCSGLTNRDCGCEVGVWLDGACVDVDECA 236  
 Db 192 CGQGLGYFAERNASHLVCSACFGPCARCSGPEESNCKLQCKGWLHLLHLCVDECGT 251  
 Qy 237 EPPPCSAQFCKNANGSYTCEDSCSCVGTGEGPGNCKECISGYAREHGQADVDGSL 296  
 Db 252 EGANGGADQGVNTEGSEYECRDCAKACLGCMGAGRGCKKCSPGYQGVGSKLDVDECE- 310  
 Qy 297 AKTCVKNENYNTPGSVVCVCPDGFETEDACVPPAEATEGESPTQLPSREDL 353  
 Db 311 -TEVPCGENKQENTEGGYRCICAEGYKQMGICV-----KEQIPGAFPIITDL 358

RESULT 5

Qy 091XD7 PRELIMINARY; PRT; 420 AA.  
 AC Q91XD7 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DE UNKNOWN (PROTEIN FOR MGC:18896).  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC Strausberg R.;  
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC010804; AA010804.1; -  
 SQ SEQUENCE 420 AA; 45717 MW; 4066BF2D739D3179 CRC64;

Query Match 48.3%; Score 968.5; DB 11; Length 420;

Best Local Similarity 47.0%; Pred. No. 6.4e-85;  
 Matches 171; Conservative 43; Mismatches 121; Indels 29; Gaps 5;

Qy 3 LPRRAALGLPLLLL-----PPAPEAAKKPTCHRCRGLVDKFNQGM 45  
 Db 4 LPPR---GLVPSLLWCLSLFLSLPVPVLPSPPPHSPRAEPHCPHCTCRALVDNFKGL 60  
 Qy 46 VDTAKNFGGNTAWSEKTLKYSSEIRLEILEGLGESSDFECNMLEAEHEHLEAW 105  
 Db 61 ERTIRDNFGGNTAWSEKTLKYSSEIRLEILEGLGESSDFECNMLEAEHEHLEAW 120  
 Qy 106 LQLKSEYDLPFVFWFCVTKLVCCSPGTGPDCLACOGSGRSCGNGHSCGSDGSG 165  
 Db 121 FHRQAEAPDLFQWLCSDSLKCCPAGTGPSCPCPGGTERPCGGYGOCEGEGTRGGSGH 180  
 Qy 166 CRCHMGYQGPLCTCDMDGYFSSLRNETHSITCTACDESKTCSGLTNRDCGCEVGVW 225  
 Db 181 CDCQAGYGEACGCGGLGYFEARENSHLVCSACFGPCARCTGPEESHCLOKKGWALH 240  
 Qy 226 GACVDVDECAEAPPPCSAAQFCKNANGSYTCEDSCSCVGTGEGPGNCKECISGYAREH 285  
 Db 241 LKCVDIDECGTEQATCGADQFCVNTGSEYECRDCAKACLGCMGAGRGCKKCSGYQGV 300